IN THE CLAIMS:

The text of all pending claims, (including withdrawn claims) is set forth below. Cancelled and not entered claims are indicated with claim number and status only. The claims as listed below show added text with underlining and deleted text with ~~strikethrough~~. The status of each claim is indicated with one of (original), (currently amended), (cancelled), (withdrawn), (new), (previously presented), or (not entered).

Please AMEND the claims as follows:

1.   (CURRENTLY AMENDED) An apparatus for extracting information from a formatted document, comprising:

an input unit for inputting a formatted document;

a unit for analyzing the input formatted document and saving analysis results containing particular typographic information;

a unit for identifying special character strings on the basis of the analysis ~~result by means of~~ results via preset values of the typographic information ~~such as font size, character font, color, etc.,~~ ;

a unit for extracting the identified special character strings; and

an output unit for outputting the extracted character strings.

2.   (CURRENTLY AMENDED) The apparatus for extracting information from a formatted document according to claim 1, wherein said unit for identifying ~~special~~ special character strings determines a certain character string as a special one on the basis of the typographic information of said formatted document when the typographic information of said character string is determined ~~as a~~ to be special typographic information.

3.   (CURRENTLY AMENDED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings identifies a certain character string as a special one on the basis of the ~~analyzing~~ analysis results with respect to said HTML document when the font size of said character string is determined to be the biggest one among the surrounding character strings.

4.   (CURRENTLY AMENDED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML

3

document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the ~~analyzing~~ analysis results with respect to said HTML document when the color and the font of said character string is determined to be a special one among the surrounding character strings.

5.   (CURRENTLY AMENDED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the ~~analyzing~~ analysis results with respect to said HTML document when the font of said character string is determined to be different from the surrounding character strings and the font of said character string ~~to be~~ is boldface.

6.   (CURRENTLY AMENDED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the ~~analyzing~~ analysis results with respect to said HTML document when the color of said character string is determined to be different from the surrounding character strings and the font of said character string ~~to be~~ is boldface.

7.   (CURRENTLY AMENDED) A method for extracting information from a formatted document ~~, comprising the following steps;~~ comprising:
        inputting a formatted document, analyzing the input formatted document and saving ~~the particular~~ analysis results containing particular typographic information;
        identifying special character strings on the basis of the analysis ~~result by means of~~ results via preset values of the typographic information ~~such as font size, character font, color, etc.~~ ;
        extracting the identified special character strings; and
        outputting the extracted character strings.

8.   (CURRENTLY AMENDED) The method according to ~~claim 8~~ claim 7, wherein in said identifying of special character strings, ~~of special character strings the step of identifying special character string,~~ a certain character string is determined as a special one on the basis of the typographic information of said formatted document when the typographic information of said character string is determined ~~as a~~ to be special typographic information.

9.   (CURRENTLY AMENDED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, the step of identifying special character string, a certain character string is determined as a special one on the basis of the analyzing analysis results with respect to said HTML document when the font size of said character string is determined to be the biggest one among the surrounding character strings.

10.   (CURRENTLY AMENDED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, the step of identifying special character string, a certain character string is determined as a special one on the basis of the analyzing results with respect to said HTML document when the color and the font of said character string is determined to be a special one among the surrounding character strings.

11.   (CURRENTLY AMENDED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, the step of identifying special character string, a certain character string is determined as a special one on the basis of the analyzing analysis results with respect to said HTML document when the font of said character string is determined to be different from the surrounding character strings and the font of said character string to be is boldface.

12.   (CURRENTLY AMENDED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, the step of identifying special character string, a certain character string is determined as a special one on the basis of the analyzing analysis results with respect to said HTML document when the color of said character string is determined to be different from the surrounding character strings and the font of said character string to be is boldface.

13.   (NEW) The apparatus for extracting information from a formatted document according to claim 1, wherein the unit for identifying special character strings on the basis of the analysis results sends the typographic information to the unit for extracting the identified special character strings if the typographic information of said character strings is beyond a range of the preset values.

14.   (NEW) The method according to claim 7, wherein said extracting extracts the special character strings if the typographic information of said character strings is beyond a range of the preset values.